Received: 22 March 2024 | Revised: 16 May 2024 | Accepted: 24 May 2024

DOI: 10.1002/glr2.12091

Grassland Research

RESEARCH ARTICLE

Chromosome-scale genome assembly of the autoalloenneaploid *Arundo donax*

Mengmeng Ren[1] | **Fupeng Liu**[1] | **Xiaohong Han**[1] | **Daohong Wu**[1] | **Hai Peng**[2]

[1]Wuhan Rundo Biotechnology Co. Ltd., Wuhan, Hubei, China

[2]Institute for Systems Biology, Jianghan University, Wuhan, Hubei, China

Correspondence
Hai Peng, Institute for Systems Biology, Jianghan University, Wuhan 430056, Hubei, China.
Email: penghai138@163.com

Handling Editor: Charlie Brummer

Funding information
Jianghan University, Grant/Award Number: 2023XKZ024; National Forestry and Grassland Administration, Grant/Award Number: KJZXXP202314

Abstract

Background: *Arundo donax* L. has great potential as an energy crop due to its high biomass yield and broad adaptability, while lack of a reference genome is a hindrance to genetic improvement efforts for this species.

Methods: Genome assembly of *A. donax* was conducted by utilizing PacBio SMRT sequencing and high-throughput chromosome conformation capture technology, with further analysis exploring the plant's ploidy, whole-genome duplication event, and evolutionary history through comparative genomics.

Results: The genome assembly of *A. donax* consists of 1.30 Gb with a contig N50 size of 33.15 Mb. A total of 74 403 gene models were identified, with over 90% of genes being functionally annotated. Karyotype and synteny analyses revealed that *A. donax* is an autoalloenneaploid ($3n = 9x = 108$) and has experienced significant gene family expansion and two whole-genome duplication events during its evolutionary history. Furthermore, utilizing the genome assembly, a variety of salinity stress-related genes were uncovered through the analysis of public RNA-seq data.

Conclusions: This study presents the initial chromosome-scale genome assembly of *A. donax*, which will advance genetic comprehension and support the genetic enhancement of this important energy crop.

KEYWORDS

Arundo donax, autoallopolyploid, comparative genomics, genome assembly

INTRODUCTION

Arundo donax L. is a tall perennial grass commonly known as giant reed. *A. donax* is a widely utilized plant species with various applications in bioenergy, papermaking, and even as an ornamental lawn grass due to its fast growth and high biomass production (Corno et al., 2014; Jámbor & Török, 2019; Pilu et al., 2013; Zhang et al., 2008). One of the significant characteristics of *A. donax* is its high productivity and fast growth rate. The plant matures rapidly and can produce large quantities of biomass in a short time, making it an attractive crop for bioenergy production (Angelini et al., 2005, 2009; Nassi o Di Nasso et al., 2013). Additionally, *A. donax* is highly tolerant of a broad range of environmental conditions, including drought, excess water, and high salinity (Nackley & Kim, 2015; Sánchez et al., 2015). These attributes make this plant an excellent candidate for sustainable agriculture and bioenergy development.

A. donax is also known for its environmental and ecological benefits. It is considered to be an excellent tool for phytoremediation, which could be used to clean up contaminated soil and water (Mirza et al., 2010, 2011). *A. donax* is effective in removing heavy metals and other pollutants from contaminated sites, making it a valuable resource in environmental restoration (Papazoglou et al., 2005). In recent years, there has been increased interest in the study and utilization of *A. donax*, which has led to considerable research efforts aimed at understanding its biological properties.

The sterile nature limits the genetic variability and delays the genetic improvement of *A. donax* (Ahmad et al., 2008; Danelli et al., 2020; Malone et al., 2017; Tarin et al., 2013). However, fundamental research into the *A. donax* genome has encountered many difficulties. The ploidy of *A. donax* is still unclear because of the high chromosome number and small size of each chromosome (Bucci et al., 2013; Mariani et al., 2010). As in another

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Grassland Research* published by John Wiley & Sons Australia, Ltd on behalf of Chinese Grassland Society and Lanzhou University.

aquatic plant, *Phragmites australis* (Clevering & Lissner, 2000), *A. donax* shows elasticity of chromosome number with variable ploidy. Different chromosome numbers in *A. donax* have been reported by several authors, including 110 (Hunter, 1934; Pizzolongo, 1962), 108 (Christopher & Abraham, 1971), and 84 chromosomes (Haddadchi et al., 2013). The complexity of the *A. donax* genome hinders further understanding and genetic improvement of this valuable energy crop.

The large and complex nature of plant genomes, coupled with varying degrees of ploidy, has long posed challenges for genome assembly in plants. Fortunately, advancements in sequencing technologies like third-generation long-read sequencing, particularly high fidelity (HiFi) reads sequencing, along with new assembly algorithms serve as the foundation for rapid advancements in plant genome sequencing (Wang et al., 2023). HiFi sequencing allows for the generation of long-read, high-quality data that can be used to accurately assemble and annotate complex genomes. It has already been proven to be a valuable tool in decoding complex plant genomes, such as allohexaploid oat (Peng et al., 2022), tetraploid potato (Sun, Jiao, Krause, et al., 2022), and autotetraploid alfalfa (Chen et al., 2020).

Here, we assembled the first chromosome-scale *A. donax* genome by combining PacBio SMRT sequencing and high-throughput chromosome conformation capture (Hi-C) technology. This assembly will facilitate better understanding of *A. donax* genomics and better utilization of *A. donax* resources.

MATERIALS AND METHODS

Plant materials

The *A. donax* plants used in this study were collected in the wild in Wuhan, Hubei, China (114.276° E, 30.789° N). Stem sections (approximately 3 cm in length) containing an axillary bud were collected and held in 70% ethanol for 1 min, surface-sterilized in a 10% (v/v) sodium hypochlorite solution (1% available chlorine) for 20 min, and rinsed twice in sterile distilled water. The sterilized section was then aseptically placed onto proliferation medium supplemented with 2 mg L^{-1} 6-enzyladenine and 0.2 mg L^{-1} naphthaleneacetic acid (NAA), and incubated under continuous fluorescent light (40 µmol m^{-2} s^{-1}) at 25°C for 45 days. Plantlets were transferred to the root medium (supplemented with NAA 0.5 mg L^{-1} + sugar 30 g L^{-1} + agar 4.5 g L^{-1}, pH 5.8) for another 45 days. Rooted tillers were then transplanted in seedling trays with nutrient soil in a greenhouse. After 45 days of growth, the seedlings were transplanted in the field at a density of 1 m × 0.5 m in April 2019. Conventional field management including weeding, irrigation, and fertilization was performed. Plants were harvested annually in December. The plants used in this study were collected in October 2022.

All media used above were based on the MS medium (Murashige & Skoog, 1962) containing 30 g L^{-1} sucrose, adjusted to pH 5.8, and solidified with 7 g L^{-1} agar.

Karyotype analysis

Karyotype analysis was conducted following a previously reported method (Bayani & Squire, 2004; Jiang, 2019). Briefly, root tips of tissue culture plantlets (about 2 cm in length) were placed in nitrous oxide (1.0 MPa) for 2 h and fixed with prechilled glacial acetic acid. Then, root tips were digested with cellulase and pectinase at 37°C for 1 h and crushed using a dissecting needle. Cell suspension was observed under an optical microscope to obtain suitable chromosome specimens. Fluorescent in situ hybridization probes were prepared using the nick translation method. Probes for telomere, 5S ribosomal DNA (rDNA) and 18S rDNA are fluorescently labeled $(TTTAGGG)_6$ oligonucleotides, pTa794 and pTa71 plasmids, respectively. After hybridization, slides were photographed under a fluorescence microscope.

DNA extraction, PacBio library preparation, and sequencing

For DNA preparation, young leaves were collected from about 50 tissue culture plantlets. High-molecular-weight genomic DNA was prepared using the QIAGEN® Genomic Kit (13343) according to the manufacturer's instructions. SMRTbell libraries were constructed followed by PacBio's standard protocol. Briefly, the quality-qualified DNA was sheared by g-TUBEs (Covaris), damage-repaired, and 20–50 kb DNA fragments were screened by the BluePippin (Sage Science) and purified using AMPure PB beads (Agilent Technologies). Sequencing was performed on a PacBio Sequel instrument with Sequel II Sequencing Kit 2.0. Raw sequencing data were filtered through SMRTlink 8.0.

MGISEQ-2000 library preparation and sequencing

Genomic DNA was extracted using the CTAB method and subjected to sonication by Covaris. DNA fragments ranging from 200 to 400 bp were selected using Agencourt AMPure XP-Medium kits. After end-repair, 3′ adenylation, adapter-ligation, and PCR amplification, the libraries were prepared using the Hieff NGS® Fast-Pace DNA Cyclization Kit (Yeasen, 13341ES96) and sequencing on the MGISEQ-2000 platform.

To facilitate the annotation of the genome, we performed messenger RNA (RNA) sequencing on four plant tissues including leaf, root, panicle, and lateral bud. Leaf and root samples were collected from about 50 tissue culture plantlets, while panicle and lateral bud samples were collected from 10 plants in the field. We purified up to 400 µg of total RNA per sample using the TRIzol-based method (Invitrogen), and subsequently treated it with DNase I. PolyA mRNA was enriched using the protocol of the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs #7490 S, USA). RNA sequencing libraries were then prepared using the Hieff NGS® Ultima Dual-mode mRNA Library Prep Kit (Yeasen, 12309ES96) following

the manufacturer's instructions. The RNA library concentration was determined using the Qubit® 3.0 Fluorometer, and the libraries were then subjected to 150 bp paired-end sequencing on an MGISEQ-2000 platform. In total, we generated 55.8 Gb raw data for the four tissue samples.

Estimation of genome size and heterozygosity

The genome size of *A. donax* was estimated based on k-mer analysis using the KMC3 program (Kokot et al., 2017). GCE and FindGSE software were used to analyze the 17-mer frequency distribution and perform genome size estimation (Sun et al., 2018). We estimated the genome heterozygosity by combining simulated data with different heterozygosity in the *Arabidopsis* genome and the distribution of 17 k-mer data of *A. donax*.

De novo assembly

The quality-filtered HiFi reads were used for genome assembly with hifiasm (v.0.12) (Cheng et al., 2021) to obtain the preliminary assembly. The completeness of genome was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) (v.4.0.5) (Simão et al., 2015) and CEGMA (v.2) (Parra et al., 2007). To assess the accuracy of the assembly, all the MGISEQ-2000 reads were mapped back to the genome using BWA (Li & Durbin, 2010), and SAMtools (Li et al., 2009) and bcftools (Danecek & McCarthy, 2017) were used to calculate the error rate of genome assembly. Also, the LTR Assembly Index (LAI) was calculated to assess the assembly quality (Ou et al., 2018).

Hi-C scaffolding

To attach the scaffolds to chromosomes, we extracted genomic DNA using young leaves from about 50 tissue culture plantlets for the Hi-C library, and the sequencing data were obtained using the MGISEQ-2000 platform. To elaborate, we first cut freshly harvested leaves into 2 cm pieces and vacuum-infiltrated them with nuclei isolation buffer supplemented with 2% formaldehyde. Crosslinking was stopped by adding glycine and performing additional vacuum infiltration. We then converted the fixed tissue into powder and resuspended it in nuclei isolation buffer to obtain a nuclei suspension. *Dpn*II was used to digest the purified nuclei with 100 U, and after marking with biotin–14-dCTP, biotin–14-dCTP from nonligated DNA ends was removed through the exonuclease activity of T4 DNA polymerase. The ligated DNA was sheared into 300–600 bp fragments, followed by blunt-end repair and A-tailing before purification through biotin–streptavidin-mediated pulldown. Finally, the Hi-C libraries were quantified and sequenced using the MGISEQ-2000 platform.

In total, 375 Gb paired-end reads were generated for Hi-C. To ensure the quality of the Hi-C raw data, we used HiC-Pro (v.2.8.1), a previously established quality control tool (Servant et al., 2015). We first filtered out low-quality sequences (quality scores < 20), adapter sequences, and sequences shorter than 30 bp using fastp (Chen et al., 2018). The clean paired-end reads were then mapped to the draft assembled sequence using bowtie2 (v.2.3.2) with the parameters "-end-to-end, –very-sensitive -L 30" to obtain the unique mapped paired-end reads (Langmead & Salzberg, 2012). We identified and retained valid interaction paired reads using HiC-Pro from the unique mapped paired-end reads for further analysis. Additionally, we filtered out invalid read pairs, including dangling-end, self-cycle, religation, and dumped products, using HiC-Pro. We further clustered, ordered, and oriented the scaffolds onto chromosomes using LACHESIS (Burton et al., 2013), with parameters CLUSTER_MIN_RE_SITES = 100, CLUSTER_MAX_LINK_DENSITY = 2.5, CLUSTER NONINFORMATIVE RATIO = 1.4, ORDER MIN N RES IN TRUNK = 60, and ORDER MIN N RES IN SHREDS = 60. Finally, we manually adjusted any placement and orientation errors showing obvious discrete chromatin interaction patterns.

Repeat element annotation

To identify tandem repeats and transposable elements (TEs) in the genome, we used GMATA (Wang & Wang, 2016) and Tandem Repeats Finder (TRF) software (Benson, 1999). GMATA identified simple repeat sequences (SSRs), while TRF recognized all tandem repeat elements in the genome. To identify TEs, we used a combination of ab initio and homology-based methods. An ab initio repeat library for the genome was predicted using MITE-hunter (Han & Wessler, 2010) and RepeatModeler with default parameters, including LTR_FINDER (Xu & Wang, 2010), LTRharvest (Ellinghaus et al., 2008), and LTR_retriever (Ou & Jiang, 2018) for plant genome. The resulting library was aligned to TEclass Repbase (http://www.girinst.org/repbase) to classify each repeat family's type. To identify repeats throughout the genome, we applied RepeatMasker (Chen, 2004) to search for known and novel TEs using the de novo repeat library and the Repbase TE library. Finally, we collated and combined overlapping TEs belonging to the same repeat class.

Gene prediction

To predict genes in a repeat-masked genome, we used three distinct approaches: ab initio prediction, homology search, and reference-guided transcriptome assembly. Homolog prediction was performed using GeMoMa by aligning homologous peptides from related species to the assembly to obtain gene structure information (Keilwagen et al., 2016). For RNA-seq-based gene prediction, we aligned filtered MGISEQ-2000 reads to the reference genome using STAR (Dobin et al., 2013) (default settings) and then assembled the resulting transcripts with stringtie2 (Kovaka et al., 2019). Open reading frames (ORFs) were predicted using PASA (Haas et al., 2008). Ab initio gene prediction was performed with Augustus (Mario & Burkhard, 2005), utilizing the default parameters with the

training set. We then used EVidenceModeler (Haas et al., 2008) to produce an integrated gene set, removing genes with TE using the TransposonPSI package (http://transposonpsi.sourceforge.net/) and filtering miscoded genes. Untranslated regions (UTRs) and alternative splicing regions were identified using PASA based on RNA-seq assembly. We retained the longest transcripts for each locus, while regions outside the ORFs were designated UTRs.

Functional annotation of gene models

Gene function information, as well as motifs and domains of their proteins, were assigned by comparing them with various publicly available databases, including SwissProt, Non-Redundant Protein Database (NR), Kyoto Encyclopedia of Gene and Genomes (KEGG), Eukaryotic Orthologous Groups of protein (KOG), and Gene Ontology (GO). We utilized the InterProScan program with default parameters to identify the putative domains and GO terms of the genes (Zdobnov & Apweiler, 2001). We also used BLASTp to compare the EvidenceModeler-integrated protein sequences against four well-known public protein databases, using an *E* value cutoff of 1e−05. We selected the results with the lowest *E* value. Results from the searches of these five databases were then concatenated to obtain a comprehensive gene function annotation.

Annotation of noncoding RNAs (ncRNAs)

For the identification of ncRNA, we utilized two main strategies: searching against databases and prediction with models. To predict transfer RNAs (tRNAs), tRNAscan-SE was applied with eukaryote parameters (Lowe & Eddy, 1997), and Infernal cmscan (Nawrocki & Eddy, 2013) was used to search the Rfam database for microRNA, rRNA, small nuclear RNA, and small nucleolar RNA detection. Additionally, we utilized RNAmmer to predict the rRNAs and their subunits (Lagesen et al., 2007).

Comparative genomic analysis

Collinearity analysis was carried out using MCScan (Python version) software. Dot plots for genome pairwise synteny were visualized with the command "python -m jcvi.graphics.dotplot."

To investigate the evolutionary history of *A. donax*, six grass species, *Oryza sativa* L., *Zea mays* L., *Sorghum bicolor* (L.) Moench., *Brachypodium distachyon* (L.) P. Beauv., *Setaria italica* L., and *Saccharum spontaneum* L., and one dicot plant, *Arabidopsis thaliana* (L.) Heynh., were used for orthologous analysis, phylogenetic analysis, and gene family expansion, and contraction analysis using OrthoVenn3 (Sun et al., 2023) with default parameters.

Whole-genome duplication (WGD) analysis was performed using wgdi (Sun, Jiao, Yang, et al., 2022). The synonymous substitution values (Ks) of syntenic gene pairs were calculated with default parameters and visualized using the R program.

RNA-seq analysis

The raw RNA-seq sequencing data (GSE121552, GSE125104) were downloaded from NCBI. We used fastp to filter the low-quality sequences, adapter sequences, and short sequences. The filtered reads were mapped to the reference genome using hisat2 (Kim et al., 2019) to obtain sam files, followed by the use of SAMtools to obtain the sorted bam file. The expression matrix was calculated using featureCounts (Liao et al., 2014). DESeq2 (Love et al., 2014) was used to identify differentially expressed genes (false discovery rate < 0.05 and $\log_2$(fold change) > 2). Gene enrichment analysis was performed using TBtools-II (Chen et al., 2023) and visualized using R program.

RESULTS

Karyotype and k-mer analysis of *A. donax*

Chromosome section and DAPI (4′,6-diamidino-2-phenylindole) staining in situ hybridization of telomere repeat sequences showed that this plant has 108 chromosomes (Figure 1a,b), consistent with a previous report (Christopher & Abraham, 1971). The rDNA fluorescence in situ hybridization revealed that 6/6 chromosomes showed strong hybridization signals of 5S rDNA and 18S rDNA (Figure 1c). To clarify the ploidy feature and genome size of *A. donax*, we conducted k-mer analysis using next-generation sequencing reads (MGI-2000 platform). The 17-mer frequency distribution curve showed two obvious peaks on 11.4 and 34.2 depth. The triple relationship of the two peaks hinted that *A. donax* is a triploid (Figure 1d), consistent with a previous assertion (Jike et al., 2020). Furthermore, the estimated genome size of *A. donax* was 1.46 Gb and genome heterozygosity was 0.8%.

High-quality genome assembly of *A. donax*

The long reads used for genome assembly were generated by PacBio platform. After quality control, we obtained 152.30 Gb reads with N50 16.92 kb. The preliminary assembly of *A. donax* was 2.24 Gb, much larger than 1.46 Gb. Considering the relatively high heterozygosity of the genome (0.8%), we used Purge_Dups (-f 0.9) (Guan et al., 2020) to obtain the 1.41 Gb nonredundant genome, occupying 96.6% of the estimated genome (1.46 Gb). GC depth analysis showed an improvement of genome quality for *A. donax* after redundancy reduction (Supporting Information S1: Figure S1).

We used BUSCO and CEGMA to assess the completeness of the genome assembly. The percentage of complete BUSCOs was 99.57% in the assembly (Supporting Information S1: Figure S2a), and the percentage of complete core genes was 92.74% (Supporting Information S1: Figure S2b), confirming the completeness of the genome assembly. Besides, we mapped the MGI short reads back to the assembly; the alignment rate was 99.60%, proving the accuracy of the genome assembly.

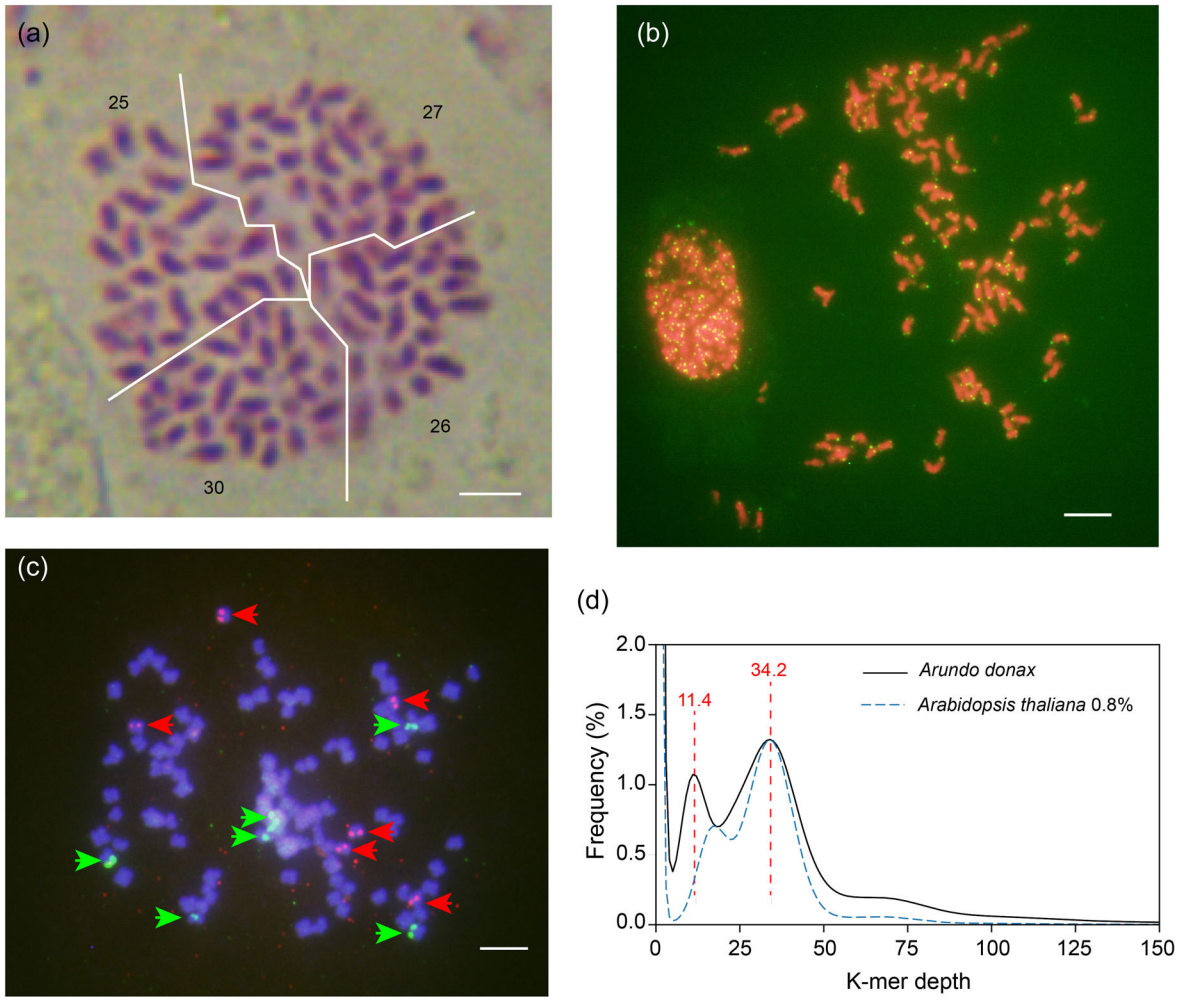


FIGURE 1 Karyotype and genome survey of *Arundo donax*. (a) DAPI (4′,6-diamidino-2-phenylindole) staining of *A. donax* chromosomes. (b) In situ hybridization of telomere repeat sequences. (c) Ribosomal DNA (rDNA) fluorescence in situ hybridization of chromosomes. Red and green arrows denote the 5S rDNA and 18S rDNA hybridization signals. Bar = 50 μm. (d) k-mer distribution curve and heterozygosity simulation curve. The horizontal axis shows the k-mer depth; the vertical axis shows the frequency of the k-mer depth. The vertical dotted lines and red numbers mark the two peaks of the k-mer distribution curve in *A. donax* genome. The genome of *Arabidopsis thaliana* was used to estimate the genome heterozygosity of *A. donax*.

As mentioned above, *A. donax* is predicted to be a triploid; thus, the haploid chromosome number is 36. Therefore, we performed Hi-C scaffolding with $n = 36$, and about 99.78% of total sequences were anchored into 36 pseudo chromosomes with sizes ranging from 18.58 to 55.91 Mb (Figure 2a). The final genome assembly is 1.30 Gb with scaffold N50 37.31 Mb (Table 1 and Figure 2b). The LAI score was 12.63, reaching the standard of reference quality. Overall, the genome assembly is a haploid genome with high quality.

Repeat elements analysis and gene model prediction

We first analyzed the interspersed repeats in *A. donax* genomes. The total length of interspersed repeats is 711.24 Mb (54.71% of the genome). Specifically, a total of 70 102 (0.07% of the genome) SSR, 89 378 (0.71% of the genome) tandem repeat sequences, and 1 400 366 (51.54% of the genome) TEs were identified in the genome. The detailed statistics of TEs are listed in Supporting Information S2 (TE).

We performed gene structure prediction by combining transcriptome prediction, homologous protein prediction, and ab initio prediction. First, we found that the alignment rates of RNA-seq data to the genome in four tissues (leaf, root, panicle, and bud) were all over 90% (Supporting Information S3), and the alignment rate of PacBio Isoseq to the genome was 99.86%, further confirming the accuracy of transcriptome data and the genome assembly. The RNA-seq and PacBio transcripts were used for gene prediction, resulting in 49 524 predicted genes. Second, we selected five Poaceae species, *S. spontaneum*, *S. bicolor*, *Z. mays*, *Triticum aestivum,* and *O. sativa,* for homologous protein prediction. A total of 94 613 genes were predicted. Third, we performed ab initio prediction, and 78 102 gene models were predicted. The final gene set was obtained by integrating the above results. In total, there were 74 403 gene models with average gene length 3507 bp, average coding sequence (CDS) length 1193 bp, average exon length 225 bp, and average intron length 538 bp (Supporting Information S2: Gene prediction). In addition to the gene model, we also predicted ncRNA. In total, 2320 rRNA, 3118 small RNA, and 1392 tRNA were identified in the genome. The parameters of the assembly are listed in Supporting Information S2 (ncRNA).

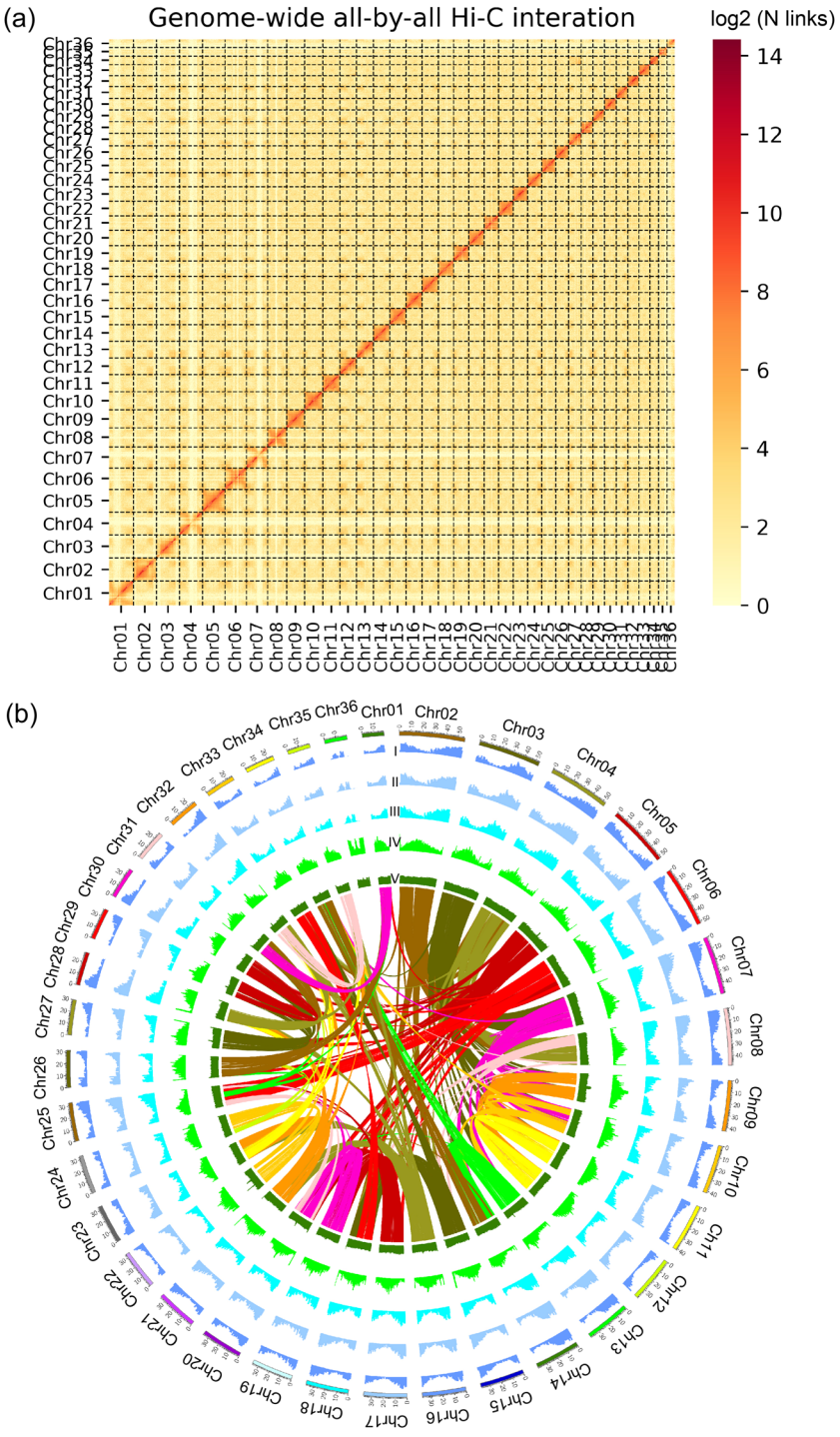


FIGURE 2 Overview of the *Arundo donax* genome assembly. (a) Genome-wide Hi-C map of *A. donax*. (b) Circos plots of *A. donax* genome assembly. The density of GC content (I), repeated sequence distribution (II), gene density (III), functional genes (IV), and expressed genes (V) were calculated using a 100 kb nonoverlap window. The innermost lines represent interchromosomal synteny.

Gene function annotation and evaluation of genome annotation

We predicted the gene function based on five databases, including the NR (Supporting Information S1: Figure S3), KEGG (Supporting Information S1: Figure S4), KOG (Supporting Information S1: Figure S5), GO (Supporting Information S1: Figure S6), and Swissprot. In total, 67 377 genes were annotated, accounting for 90.56% of all the predicted genes (Supporting Information S2: Gene prediction). A total of 16 877 genes were co-annotated among these five databases (Supporting Information S1: Figure S7).

The annotated gene sets were evaluated using BUSCO. Among the 1614 BUSCO groups, about 98.45% of complete gene elements can be found in the annotated gene set, indicating that the majority of conservative gene predictions are relatively complete and confirming the high reliability of the gene prediction result. Besides, the proportion of expressed genes in four tissues ranged from 71.09% to 80.34%. Total expressed genes account for 86.60% of the whole gene sets (Supporting Information S2: Transcripts). The gene structure in *A. donax* genome showed similar distribution trends as other Poaceae plants, including gene length, CDS length, exon length, exon number, intron

TABLE 1 Statistics of the genome assembly.

	Statistics
Assembly features	
Number of scaffolds	65
Total size of scaffolds	1299.92 Mb
Longest scaffold	55.91 Mb
Shortest scaffold	18.58 Mb
Mean scaffold size	20.00 Mb
N50 scaffold length	37.31 Mb
L50 scaffold count	7
Scaffold GC content	44.07%
Scaffold N content	0.0002%
Percentage of assembly in scaffolded contigs	99.78%
Average number of contigs per scaffold	1.46
BUSCO (complete)	99.57%
LAI	12.63
Gene models	
Number of gene models	74 403
Mean coding sequence length	1192.62
Mean number of exons per gene	5.31
Mean exon length	224.79
Mean intron length	537.64
Nonprotein-coding RNA	
Number of rRNA	2320
Number of sRNA	3118
Number of regulatory	19
Number of tRNA	1392

Abbreviations: BUSCO, Benchmarking Universal Single-Copy Orthologs; LAI, LTR Assembly Index; rRNA, ribosomal RNA; sRNA, small RNA; tRNA, transfer RNA.

length, and intron number (Supporting Information S1: Figure S8), demonstrating the reliability of the genome annotations.

A. donax is an autoalloenneaploid

Based on the protein sequence of the genome assembly, we performed intragenomic comparison within *A. donax*. The discontinuous synteny chromosome segments revealed that *A. donax* has undergone multiple chromosome rearrangements during its evolutionary process. Interestingly, most single chromosome segments can be aligned to two other chromosome segments (Figure 3a). Moreover, synteny analysis of *A. donax* and *S. italica* showed a 1:3 syntenic relationship (Figure 3b and Supporting Information S1: Figure S9). Based on these results, we deduced that *A. donax* is an enneaploid.

To determine whether *A. donax* is autoenneaploid or alloenneaploid, we used SubPhaser (Jia et al., 2022) to split the subgenomes of *A. donax*. The k-mer-based heatmap showed that the chromosomes were clustered into two groups, in which subgenome A has 12 chromosomes and subgenome B has 24 chromosomes (Figure 3c). Therefore, our results jointly proved that *A. donax* is autoalloenneaploid, and the karyotype is AAABBBBBB ($3n = 9x = 108$). The WGD analysis showed that *A. donax* underwent two WGD events: one is the ancient ρ event shared by Poaceae plants (Wang et al., 2015) and another is a recent burst WGD event that occurred ~13.5 million years ago (MYA) (Figure 3d).

Gene family clustering analysis of *A. donax*

Gene family clustering with five other grass species and a dicotyledon, as listed in the MATERIALS AND METHODS section, identified a total of 21 162 gene clusters in *A. donax* genomes, in which 9063 clusters were shared by all the above species and 1989 clusters were unique to *A. donax* (Figure 4a). A total of 121 single-copy orthologs shared by *A. donax, A. thaliana* and six other grass species were used for phylogenetic analysis and divergence time estimation, which showed that *A. donax* and other grass species shared a common ancestor ~49.5 MYA (Figure 4b).

A. donax has undergone dramatic gene family expansion during its evolution (Figure 4c). The 611 expanded gene families were enriched in GO terms like "response to water deprivation," "response to oxidative stress," "response to osmotic stress," "response to cold," and "response to heat" (Figure 4d), hinting that *A. donax* emerged within the grass family because of the severe environment on earth at that time.

Salt stress response gene mining of *A. donax*

Two previous studies had identified multiple salt stress response genes using RNA-seq (Sicilia et al., 2019, 2020), while the analysis was based on transcript assembly and offered limited information of specific genes. To deeply mine salt stress response genes, we reanalyzed the public RNA-seq data using the above genome assembly. The mapping rates of the two RNA-seq data to the genome assembly in this study were 67.0%–71.1% and 71.1%–74.4% (Supporting Information S3), further confirming the accuracy of the assembly. The gene expression heatmap using transcripts per million values showed that one of the two studies showed low data consistency among the biological replicates (Supporting Information S1: Figure S10), while another study showed better data quality (Figure 5a). Therefore, we used the data with high consistency that contained two levels of salt treatment (severe and extreme) to perform the following analysis.

A total of 3471 differentially expressed genes (DEGs) were identified in three comparisons. In detail, 956 DEGs were identified in CK (control) versus severe (240 upregulated and 746 downregulated), 2875 DEGs were identified in CK versus extreme (1119 upregulated and 1756 downregulated), and 1395 DEGs were identified in severe versus extreme (607 upregulated and 787 downregulated) (Figure 5b and Supporting Information S4).

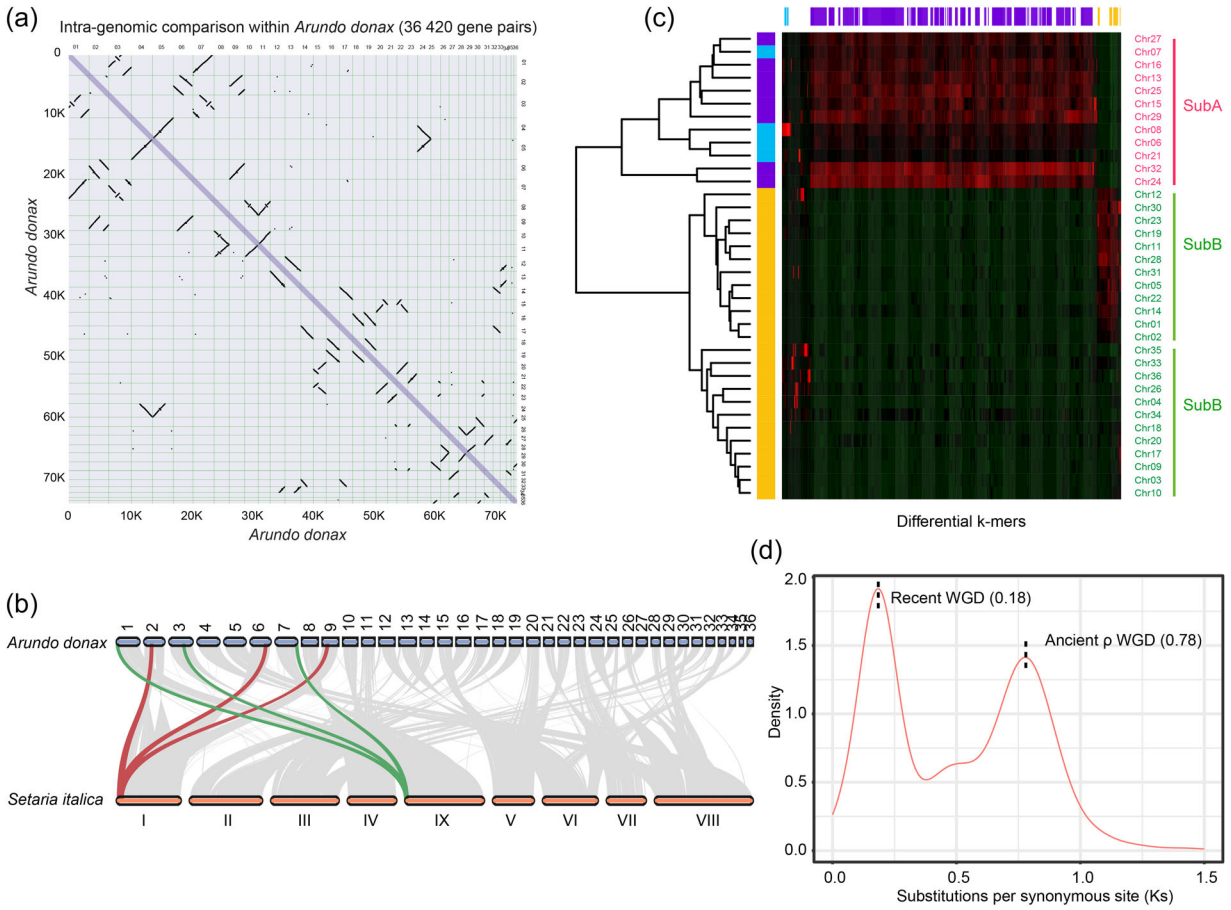


FIGURE 3 Comparative genomic analysis and whole-genome duplication events of *Arundo doanx*. (a) Dot plot illustrating the comparative analysis of *A. donax* versus *A. donax*. (b) Syntenic relationships between *A. donax* and *S. italica*. Colored stripes represent the correspondence of syntenic chromosome segments. (c) Subgenomes of *A. donax*. Clustering based on specific k-mers to partition the two subgenomes. Red and green chromosomes correspond to the A and B subgenomes, respectively. (d) Frequency distributions of synonymous substitution rates (Ks) of homologous gene pairs located in the collinearity blocks of *A. donax* versus *A. donax*. The two peaks indicate the recent WGD in *A. donax* and the shared WGD (ρ event) in the grass family.

Next, 584 DEGs (overlap of CK_severe and CK_extreme) were used for GO enrichment analysis. Interestingly, the top 15 enrichment pathways contained "response to water deprivation," "response to water," and "response to salt," confirming that these DEGs were indeed a response to salt stress (Figure 5c).

DISCUSSION

A. donax is a promising energy crop with a complex genome

The rapidly deteriorating environment on Earth has made an urgent case for carbon emission reduction (Patz et al., 2014). Using bioenergy to replace a proportion of fossil energy consumption can potentially make a substantial contribution toward reducing carbon emissions (Reid et al., 2020). However, land-intensive bioenergy is limited by the finite land resource. Especially in China, because of the huge population size, croplands are inviolable to guarantee food security. However, there are about 78 million hectares of marginal land besides the cropland (Tang et al., 2010). Due to climate, terrain, soil, and other limiting factors, marginal land cannot be planted with food crops, while energy crops can often grow on these land areas owing to their excellent environmental adaptability. Of all the energy crops, the high yield, extensive adaptability, and low inputs make *A. donax* an ideal energy crop.

Our results partly unravel the composition of the genome of *A. donax*. We generated the first chromosome-scale assembly of the *A. donax* genome using HiFi reads and Hi-C technology. Combining karyotype analysis, k-mer analysis, and Hi-C scaffolding, we inferred that *A. donax* is an autoalloenneaploid ($3n = 9x = 108$), which corroborates a recent study (Jike et al., 2020). The interchromosomal synteny analysis of *A. donax* and *S. italica* confirms that *A. donax* is an autoalloenneaploid. Based on these results, we arbitrarily inferred that the chromosome evolutionary process of *A. donax* underwent three WGD events and two whole-genome triplication (WGT) events. The first WGT resulted in an autoallohexaploid, and the second WGT finally resulted in the present autoalloenneaploid genome of *A. donax* (Supporting Information S1: Figure S11). However, WGD analysis based on Ks values only detected two WGD events (Figure 3d). We assumed that an ancient WGD event became undetectable during evolution. To further unveil the composition of the *A. donax*

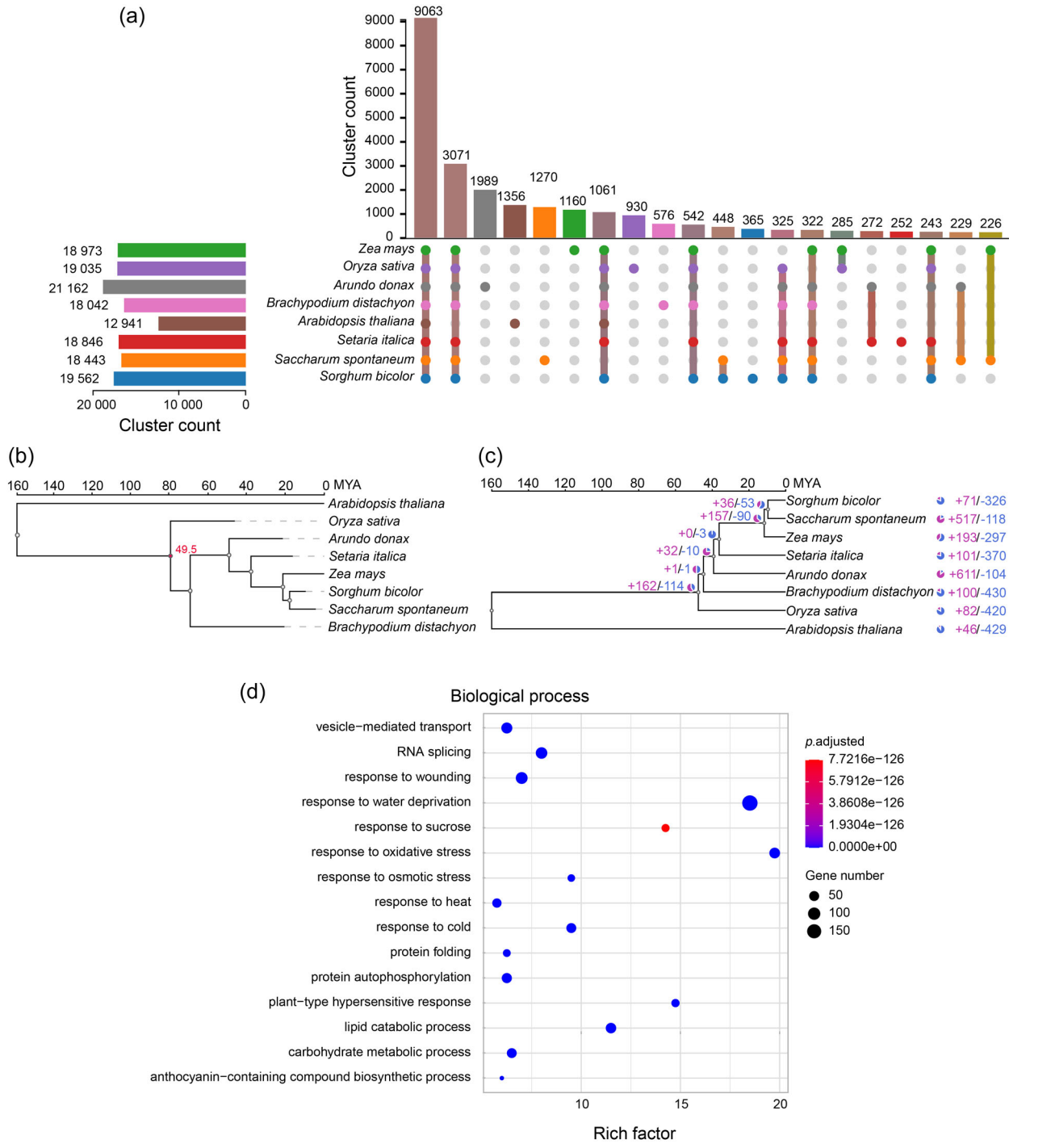


FIGURE 4 Gene family clustering analysis of *Arundo donax*. (a) The UpSet table displays the count of orthologous clusters for each species, along with the count of unique gene clusters and the count of shared gene clusters among different species. (b) Timetree of *A. donax* and several other Poaceae plants. The evolutionary relationships between species were estimated by highly conserved single-copy genes. MYA, million years ago. (c) A pie chart was utilized to visualize gene families with altered gene numbers, representing the contracted gene families (depicted in purple) and expanded gene families (depicted in blue). (d) Gene Ontology (GO) enrichment of 611 expanded gene clusters of *A. donax* (top 15 of biological process category). The color of dots denotes the adjusted *p* value using the hypergeometric test. The size of dots denotes the gene number of GO terms.

genome, more *Arundo* genus or Arundinoideae plants need to be studied.

The reference genome will contribute to genetic improvement of *A. donax*

Clonally propagated crops like potatoes often show small genetic gains. Using precise genome design, vigorous F_1 hybrid potatoes have recently been developed (Zhang et al., 2021). Like potato, *A. donax* is completely agamic and shows a narrow genetic variability. To accelerate the genetic improvement of *A. donax* in the future, we need to elucidate the mechanism of infertility of *A. donax*. In addition, the genetic transformation of *A. donax* is difficult, and no transgenic *A. donax* has been reported to date. Several developmental regulators, including BBM, WUS, GRF4, and GIF, are reported to improve transformation efficiency in crops like corn and wheat (Chen et al., 2022). Therefore, cloning key developmental regulators in *A. donax*

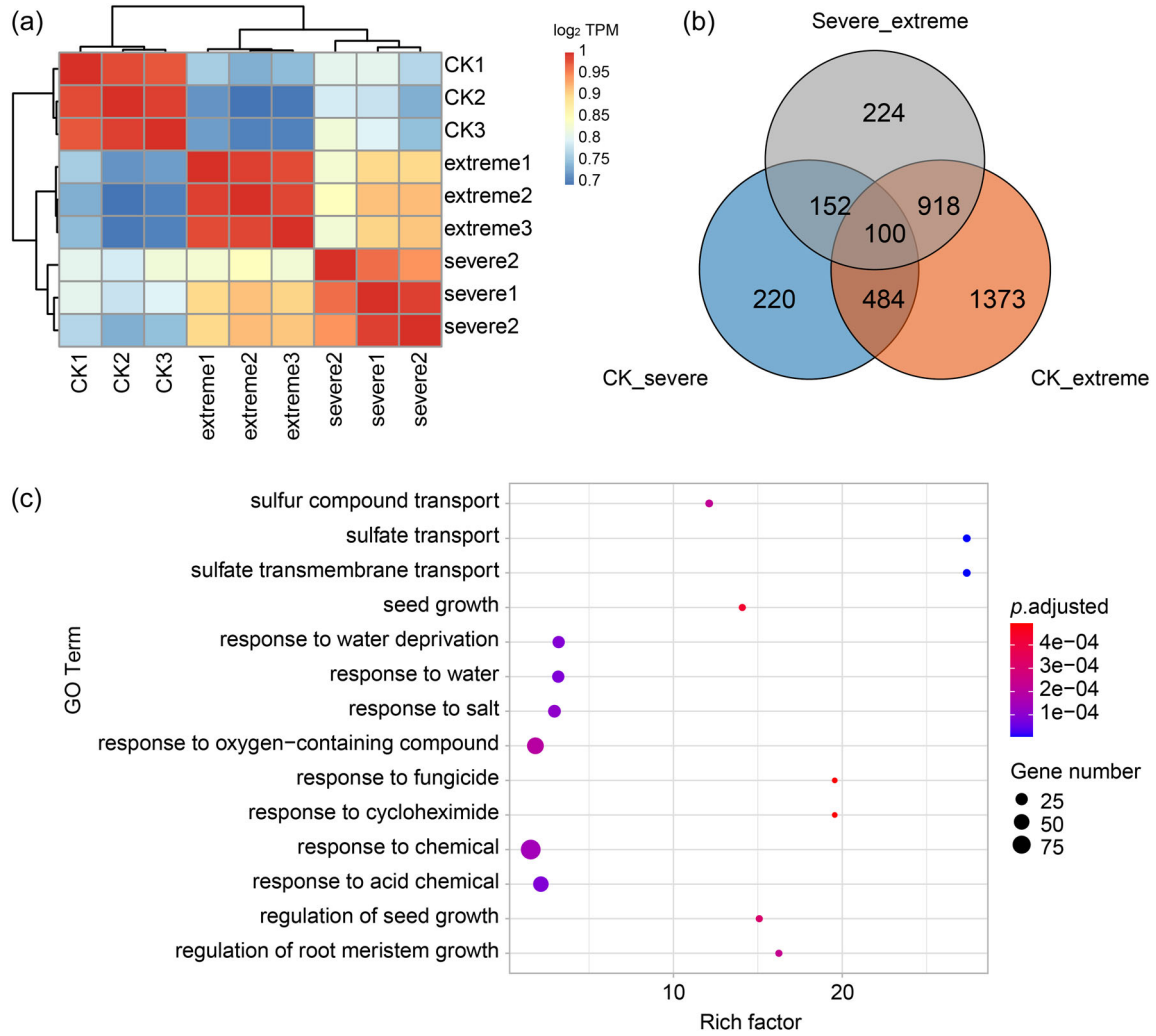


FIGURE 5 Salt stress response gene mining. (a) Gene expression heatmap of the nine samples using transcripts per million values. (b) Venn diagrams of three sets of differentially expressed genes (DEGs). (c) Gene Ontology (GO) enrichment of 584 DEGs shared by CK_severe and CK_extreme (top 15 of the biological process category). The color of the dot denotes the adjusted *p* value using the hypergeometric test. The size of the dot denotes the gene number of GO terms. CK, control.

may help to break down technical barriers in genetic transformation of *A. donax*. To overcome the above two difficulties, a reference genome is indispensable.

Until now, little was known about the *A. donax* genome, except for a few transcriptome studies (Evangelistella et al., 2017; Fu et al., 2016; Sablok et al., 2014; Sicilia et al., 2019). We reanalyzed one of the RNA-seq public data sets and mined multiple salt stress response genes. Among the DEGs, 25 genes were annotated to be responses to salt, water, and water deprivation (Figure 5c and Supporting Information S4), which showed an obvious connection with salt-stress-related pathways and are potential targets for genetic improvement. This result demonstrates that analysis of RNA-seq data with the genome assembly that we produced provided a feasible approach to mining specific functional genes and will undoubtedly contribute to the genetic improvement of *A. donax*. Future research on *A. donax* will continue to uncover new insights into the plant's biological properties and the potential for its utilization in different industries. It will undoubtedly uncover more valuable information and offer even more exciting opportunities to take advantage of this remarkable plant species.

CONCLUSIONS

Lack of a reference genome is a main obstacle to genetic improvement of *A. doanx*. In this study, we carried out the first chromosome-scale genome assembly of *A. donax* and indicated that *A. donax* is an autoalloenneaploid ($3n = 9x = 108$) through comparative genomic analysis. Based on the assembly, we mined multiple salt response genes using public RNA-seq data, proving the application prospect of this assembly in molecular breeding of *A. donax* in the future.

AUTHOR CONTRIBUTIONS

Mengmeng Ren: Data curation; software; writing—original draft. **Fupeng Liu**: Data curation. **Xiaohong Han**: Data curation. **Daohong Wu**: Resources; validation. **Hai Peng**: Funding acquisition; supervision.

ACKNOWLEDGMENTS

This work was supported by Jianghan University (2023XKZ024) and the National Forestry and Grassland Administration (KJZXXP202314).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The genome assembly and gene annotation have been deposited in the Genome Warehouse in the National Genomics Data Center (https://ngdc.cncb.ac.cn/gwh) under BioProject number PRJCA024002 and accession number GWHERLM00000000. Source data are provided within this paper.

ORCID

Hai Peng http://orcid.org/0000-0001-7827-3322

REFERENCES

Ahmad, R., Liow, P. S., Spencer, D. F., & Jasieniuk, M. (2008). Molecular evidence for a single genetic clone of invasive *Arundo donax* in the United States. *Aquatic Botany*, *88*(2), 113–120. https://doi.org/10.1016/j.aquabot.2007.08.015

Angelini, L. G., Ceccarini, L., & Bonari, E. (2005). Biomass yield and energy balance of giant reed (*Arundo donax* L.) cropped in central Italy as related to different management practices. *European Journal of Agronomy*, *22*, 375–389. https://doi.org/10.1016/j.eja.2004.05.004

Angelini, L. G., Ceccarini, L., Nasso, N., & Bonari, E. (2009). Comparison of *Arundo donax* L. and *Miscanthus* × *giganteus* in a long-term field experiment in central Italy: Analysis of productive characteristics and energy balance. *Biomass & Bioenergy*, *33*(4), 635–643. https://doi.org/10.1016/j.biombioe.2008.10.005

Bayani, J., & Squire, J. A. (2004). Fluorescence in situ hybridization (FISH). *Current Protocols in Cell Biology*, *3*(1), 22–24. https://doi.org/10.1002/0471143030.cb2204s23

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580. https://doi.org/10.1093/nar/27.2.573

Bucci, A., Cassani, E., Landoni, M., Cantaluppi, E., & Pilu, R. (2013). Analysis of chromosome number and speculations on the origin of *Arundo donax* L. (Giant Reed). *Cytology and Genetics*, *47*(4), 237–241. https://doi.org/10.3103/S0095452713040038

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, *31*, 1119–1125. https://doi.org/10.1038/nbt.2727

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., Liu, Y., Feng, J., Chen, H., He, Y., & Xia, R. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Molecular Plant*, *16*(11), 1733–1742. https://doi.org/10.1016/j.molp.2023.09.010

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F., Liu, W., Li, Y., Liu, Y., Zhang, X., Zhang, R., Zhang, Y., Li, Y., Wang, K., He, H., Wang, Z., Fan, G., Yang, H., … Qiu, Q. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature Communications*, *11*(1), 2494. https://doi.org/10.1038/s41467-020-16338-x

Chen, N. S. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, *4*, 10. https://doi.org/10.1002/0471250953.bi0410s05

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Z., Debernardi, J. M., Dubcovsky, J., & Gallavotti, A. (2022). Recent advances in crop transformation technologies. *Nature Plants*, *8*(12), 1343–1351. https://doi.org/10.1038/s41477-022-01295-8

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

Christopher, J., & Abraham, A. (1971). Studies on the cytology and phylogeny of South Indian grasses I. Subfamilies Bambusoideae, Oryzoideae, Arundinoideae and Festucoideae. *Cytologia*, *36*, 579–594. https://doi.org/10.1508/cytologia.36.579

Clevering, O. A., & Lissner, J. (2000). Erratum to "Taxonomy, chromosome numbers, clonal diversity and population dynamics of *Phragmites australis*". *Aquatic Botany*, *66*, 249–250. https://doi.org/10.1016/S0304-3770(00)00094-2

Corno, L., Pilu, R., & Adani, F. (2014). *Arundo donax* L.: A non-food crop for bioenergy and bio-compound production. *Biotechnology Advances*, *32*(8), 1535–1549. https://doi.org/10.1016/j.biotechadv.2014.10.006

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics*, *33*(13), 2037–2039. https://doi.org/10.1093/bioinformatics/btx100

Danelli, T., Laura, M., Savona, M., Landoni, M., Adani, F., & Pilu, R. (2020). Genetic improvement of *Arundo donax* L.: Opportunities and challenges. *Plants*, *9*(11), 1584. https://doi.org/10.3390/plants9111584

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, *9*, 18. https://doi.org/10.1186/1471-2105-9-18

Evangelistella, C., Valentini, A., Ludovisi, R., Firrincieli, A., Fabbrini, F., Scalabrin, S., Cattonaro, F., Morgante, M., Mugnozza, G. S., Keurentjes, J. J. B., & Harfouche, A. (2017). De novo assembly, functional annotation, and analysis of the giant reed (*Arundo donax* L.) leaf transcriptome provide tools for the development of a biofuel feedstock. *Biotechnology for Biofuels*, *10*(1), 138. https://doi.org/10.1186/s13068-017-0828-7

Fu, Y., Poli, M., Sablok, G., Wang, B., Liang, Y., La Porta, N., Velikova, V., Loreto, F., Li, M., & Varotto, C. (2016). Dissection of early transcriptional responses to water stress in *Arundo donax* L. by unigene-based RNA-seq. *Biotechnology for Biofuels*, *9*, 54. https://doi.org/10.1186/s13068-016-0471-8

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. https://doi.org/10.1101/729962

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, *9*(1),R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haddadchi, A., Gross, C. L., & Fatemi, M. (2013). The expansion of sterile *Arundo donax* (Poaceae) in southeastern Australia is accompanied by genotypic variation. *Aquatic Botany*, *104*, 153–161. https://doi.org/10.1016/j.aquabot.2012.07.006

Han, Y., & Wessler, S. R. (2010). MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Research*, *38*, e199. https://doi.org/10.1093/nar/gkq862

Hunter, A. W. S. (1934). A karyosystematic investigation in the Gramineae. *Canadian Journal of Research*, *11*(2), 213–241. https://doi.org/10.1139/cjr34-087

Jámbor, A., & Török, Á. (2019). The economics of *Arundo donax*—A systematic literature review. *Sustainability*, *11*, 4225. https://doi.org/10.3390/su11154225

Jia, K. H., Wang, Z. X., Wang, L., Li, G. Y., Zhang, W., Wang, X. L., Xu, F. J., Jiao, S. Q., Zhou, S. S., Liu, H., Ma, Y., Bi, G., Zhao, W., El-Kassaby, Y. A., Porth, I., Li, G., Zhang, R. G., & Mao, J. F. (2022). SubPhaser: A robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytologist*, *235*(2), 801–809. https://doi.org/10.1111/nph.18173

Jiang, J. (2019). Fluorescence *in situ* hybridization in plants: Recent developments and future applications. *Chromosome Research*, *27*(3), 153–165. https://doi.org/10.1007/s10577-019-09607-z

Jike, W., Li, M., Zadra, N., Barbaro, E., Sablok, G., Bertorelle, G., Rota-Stabelli, O., & Varotto, C. (2020). Phylogenomic proof of recurrent demipolyploidization and evolutionary stalling of the "Triploid Bridge" in *Arundo* (Poaceae). *International Journal of Molecular Sciences*, *21*(15), 5247. https://doi.org/10.3390/ijms21155247

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, *44*, e89. https://doi.org/10.1093/nar/gkw092

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kokot, M., Długosz, M., & Deorowicz, S. (2017). KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics*, *33*(17), 2759–2761. https://doi.org/10.1093/bioinformatics/btx304

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, *20*(1), 278. https://doi.org/10.1186/s13059-019-1910-1

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108. https://doi.org/10.1093/nar/gkm160

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964. https://doi.org/10.1093/nar/25.5.955

Malone, J. M., Virtue, J. G., Williams, C., & Preston, C. (2017). Genetic diversity of giant reed (*Arundo donax*) in Australia. *Weed Biology and Management*, *17*, 17–28. https://doi.org/10.1111/wbm.12111

Mariani, C., Cabrini, R., Danin, A., Piffanelli, P., Fricano, A., Gomarasca, S., Dicandilo, M., Grassi, F., & Soave, C. (2010). Origin, diffusion and reproduction of the giant reed (*Arundo donax* L.): A promising weedy energy crop. *Annals of Applied Biology*, *157*(2), 191–202. https://doi.org/10.1111/j.1744-7348.2010.00419.x

Mario, S., & Burkhard, M. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, *33*(Suppl. 2), W465–W467. https://doi.org/10.1093/nar/gki458

Mirza, N., Mahmood, Q., Pervez, A., Ahmad, R., Farooq, R., Shah, M. M., & Azim, M. R. (2010). Phytoremediation potential of *Arundo donax* in arsenic-contaminated synthetic wastewater. *Bioresource Technology*, *101*, 5815–5819. https://doi.org/10.1016/j.biortech.2010.03.012

Mirza, N., Pervez, A., Mahmood, Q., Shah, M. M., & Shafqat, M. N. (2011). Ecological restoration of arsenic contaminated soil by *Arundo donax* L. *Ecological Engineering*, *37*(12), 1949–1956. https://doi.org/10.1016/j.ecoleng.2011.07.006

Murashige, T., & Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiologia Plantarum*, *15*, 473–497. https://doi.org/10.1111/j.1399-3054.1962.tb08052.x

Nackley, L. L., & Kim, S. H. (2015). A salt on the bioenergy and biological invasions debate: Salinity tolerance of the invasive biomass feedstock *Arundo donax*. *Global Change Biology Bioenergy*, *7*, 752–762. https://doi.org/10.1111/gcbb.12184

Nassi o Di Nasso, N., Roncucci, N., & Bonari, E. (2013). Seasonal dynamics of aboveground and belowground biomass and nutrient accumulation and remobilization in giant reed (*Arundo donax* L.): A three-year study on marginal land. *Bioenergy Research*, *6*(2), 725–736. https://doi.org/10.1007/s12155-012-9289-9

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Ou, S., Chen, J., & Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research*, *46*(21), 126. https://doi.org/10.1093/nar/gky730

Ou, S., & Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, *176*(2), 1410–1422. https://doi.org/10.1104/pp.17.01310

Papazoglou, E. G., Karantounias, G. A., Vemmos, S. N., & Bouranis, D. L. (2005). Photosynthesis and growth responses of giant reed (*Arundo donax* L.) to the heavy metals Cd and Ni. *Environment International*, *31*(2), 243–249. https://doi.org/10.1016/j.envint.2004.09.022

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*(9), 1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Patz, J. A., Frumkin, H., Holloway, T., Vimont, D. J., & Haines, A. (2014). Climate change: Challenges and opportunities for global health. *Journal of the American Medical Association*, *312*(5), 1565–1580. https://doi.org/10.1001/jama.2014.13186

Peng, Y., Yan, H., Guo, L., Deng, C., Wang, C., Wang, Y., Kang, L., Zhou, P., Yu, K., Dong, X., Liu, X., Sun, Z., Peng, Y., Zhao, J., Deng, D., Xu, Y., Li, Y., Jiang, Q., Li, Y., … Ren, C. (2022). Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nature Genetics*, *54*(8), 1248–1258. https://doi.org/10.1038/s41588-022-01127-7

Pilu, R., Manca, A., & Landoni, M. (2013). *Arundo donax* as an energy crop: Pros and cons of the utilization of this perennial plant. *Maydica*, *58*(1), 54–59.

Pizzolongo, P. (1962). Osservazioni cariologichec su *Arundo donax* e *Arundo plinii*. *Annuali Botany*, *27*, 173–187.

Reid, W. V., Ali, M. K., & Field, C. B. (2020). The future of bioenergy. *Global Change Biology*, *26*(1), 274–286. https://doi.org/10.1111/gcb.14883

Sablok, G., Fu, Y., Bobbio, V., Laura, M., Rotino, G. L., Bagnaresi, P., Allavena, A., Velikova, V., Viola, R., Loreto, F., Li, M., & Varotto, C. (2014). Fuelling genetic and metabolic exploration of C_3 bioenergy crops through the first reference transcriptome of *Arundo donax* L. *Plant Biotechnology Journal*, *12*, 554–567. https://doi.org/10.1111/pbi.12159

Sánchez, E., Scordia, D., Lino, G., Arias, C., Cosentino, S. L., & Nogués, S. (2015). Salinity and water stress effects on biomass production in different *Arundo donax* L. clones. *Bioenergy Research*, *8*(4), 1461–1479. https://doi.org/10.1007/s12155-015-9652-8

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*, 259. https://doi.org/10.1186/s13059-015-0831-x

Sicilia, A., Santoro, D. F., Testa, G., Cosentino, S. L., & Piero, A. R. L. (2020). Transcriptional response of giant reed (*Arundo donax L.*) low ecotype to long-term salt stress by unigene-based RNAseq. *Phytochemistry*, *177*, 112436. https://doi.org/10.1016/j.phytochem.2020.112436

Sicilia, A., Testa, G., Santoro, D. F., Cosentino, S. L., & Lo Piero, A. R. (2019). RNASeq analysis of giant cane reveals the leaf transcriptome dynamics under long-term salt stress. *BMC Plant Biology*, *19*(1), 355. https://doi.org/10.1186/s12870-019-1964-y

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sun, H., Ding, J., Piednoël, M., & Schneeberger, K. (2018). FindGSE: Estimating genome size variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics*, *34*(4), 550–557. https://doi.org/10.1093/bioinformatics/btx637

Sun, H., Jiao, W. B., Krause, K., Campoy, J. A., Goel, M., Folz-Donahue, K., Kukat, C., Huettel, B., & Schneeberger, K. (2022). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics*, *54*(3), 342–348. https://doi.org/10.1038/s41588-022-01015-0

Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., & Wang, Y. (2023). OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research*, *51*(W1), W397–W403. https://doi.org/10.1093/nar/gkad313

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X., & Liu, J. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes.

Molecular Plant, *15*(12), 1841–1851. https://doi.org/10.1016/j.molp.2022.10.018

Tang, Y., Xie, J. S., & Geng, S. (2010). Marginal land-based biomass energy production in China. *Journal of Integrative Plant Biology*, *52*(1), 112–121. https://doi.org/10.1111/j.1744-7909.2010.00903.x

Tarin, D., Pepper, A. E., Goolsby, J. A., Moran, P. J., Arquieta, A. C., Kirk, A. E., & Manhart, J. R. (2013). Microsatellites uncover multiple introductions of clonal giant reed (*Arundo donax*). *Invasive Plant Science & Management*, *6*(3), 328–338. https://doi.org/10.1614/ipsm-d-12-00085.1

Wang, X., Wang, J., Jin, D., Guo, H., Lee, T. H., Liu, T., & Paterson, A. H. (2015). Genome alignment spanning major poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Molecular Plant*, *8*(6), 885–898. https://doi.org/10.1016/j.molp.2015.04.004

Wang, X., & Wang, L. (2016). GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science*, *7*(7), 1350. https://doi.org/10.3389/fpls.2016.01350

Wang, Y., Yu, J., Jiang, M., Lei, W., Zhang, X., & Tang, H. (2023). Sequencing and assembly of polyploid genomes. *Methods in Molecular Biology*, *2545*, 429–458. https://doi.org/10.1007/978-1-0716-2561-3_23

Xu, Z., & Wang, H. (2010). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, *35*(Web Server issue), W265–W268. https://doi.org/10.1093/nar/gkm286

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*(9), 847–848. https://doi.org/10.1093/bioinformatics/17.9.847

Zhang, C., Yang, Z., Tang, D., Zhu, Y., Wang, P., Li, D., Zhu, G., Xiong, X., Shang, Y., Li, C., & Huang, S. (2021). Genome design of hybrid potato. *Cell*, *184*(15), 3873–3883.e12. https://doi.org/10.1016/j.cell.2021.06.006

Zhang, J., Li, Y., Zhang, C., & Jing, Y. (2008). Adsorption of malachite green from aqueous solution onto carbon prepared from *Arundo donax* root. *Journal of Hazardous Materials*, *150*(3), 774–782. https://doi.org/10.1016/j.jhazmat.2007.05.036

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ren, M., Liu, F., Han, X., Wu, D., & Peng, H. (2024). Chromosome-scale genome assembly of the autoalloenneaploid *Arundo donax*. *Grassland Research*, *3*(3), 230–242. https://doi.org/10.1002/glr2.12091